STATE OF ILLINOIS

ILLINOIS COMMERCE COMMISSION

Illinois Bell Telephone Company, AT&T Communications of Illinois, Inc., TCG Illinois, TCG Chicago, TCG St. Louis, WorldCom, Inc., McLeodUSA Telecommunications Services, Inc., XO Illinois, Inc., NorthPoint Communications, Inc., Rhythms Netconnection and Rhythms Links, Inc., Sprint Communications L.P., Focal Communications Corporation of Illinois, And Gabriel Communications of Illinois Inc. Petition for Resolution of Disputed Issues Pursuant to Condition (30) of the SBC/Ameritech Merger Order.	)	No. 01-0120

DIRECT TESTIMONY

OF

DR. DANIEL S. LEVY

On Behalf of

AMERITECH ILLINOIS

July 13, 2001

DIRECT TESTIMONY
OF DR. DANIEL S. LEVY
ON BEHALF OF AMERITECH ILLINOIS

I. BACKGROUND, QUALIFICATIONS, AND PURPOSE OF TESTIMONY.

Q. PLEASE STATE YOUR NAME AND BUSINESS ADDRESS.

A. My name is Dr. Daniel S. Levy. My business address is 33 West Monroe Street, Chicago, Illinois.

Q. BY WHOM ARE YOU EMPLOYED AND IN WHAT CAPACITY?

A. I am a Partner at Arthur Andersen, LLP, where I serve as the National Director of Economic Consulting for Arthur Andersen's Business Consulting Group. In that capacity, I advise clients as to the use of statistical analysis and techniques in business and in judicial and regulatory proceedings.

Q. WHAT IS THE PURPOSE OF YOUR TESTIMONY?

A. In this proceeding, the Commission will decide whether Ameritech Illinois should keep its current plan for paying remedies to competing local exchange carriers ("CLECs") in the event it fails to meet certain performance standards, or whether to replace the current remedy plan with one proposed by the CLECs. Each plan has a different methodology for using statistical analysis to test compliance with performance standards. The purpose of my testimony is to explain and compare the statistical methodologies used in the Ameritech Illinois and CLEC plans, and to show why Ameritech Illinois' methodology is balanced, practical, scientifically valid, and consistent with the goals of the Telecommunications Act of 1996 ("1996 Act").

Q. PLEASE DESCRIBE YOUR BACKGROUND AND QUALIFICATIONS FOR REACHING THAT CONCLUSION.

A. I have a PhD. in Economics from The University of Chicago. I have over 20 years of experience in research and advising clients, particularly on matters related to statistics. And for more than three years, I have advised Ameritech Illinois and its affiliates with respect to the implementation and ongoing administration of performance remedy plans in general, and the remedy plan that is used in Illinois in particular.

Q. ARE YOU ALSO FAMILIAR WITH THE CLEC REMEDY PLAN?

A. Yes. I have reviewed the plan that the CLECs filed in this docket on March 12, 2001. I have also reviewed virtually identical plans that these CLECs submitted in Michigan, Indiana, Ohio, and Wisconsin, and I have reviewed testimony by CLEC witnesses regarding those plans. Further, I have reviewed similar statistical methodologies proposed by CLECs for use in the ongoing third-party tests of operations support systems ("OSS") in the Ameritech states.

Q. IN ADDITION TO YOUR WORK AT ARTHUR ANDERSEN, WHAT POSITIONS HAVE YOU HELD?

A. Prior to joining Arthur Andersen, I performed research and consulting work for Needham-Harper Worldwide Advertising, The University of Chicago Computation Center, SPSS Inc., The RAND Corporation, and Charles River Associates. I joined Arthur Andersen in 1995 as an economist. I became the Regional Director of Economics for Arthur Andersen's Central Region (which includes Illinois) one year later. I was promoted to National Director in 1998, and have served in that position since. My resume is attached hereto as Attachment A.

Q. WHAT IS THE PURPOSE OF STATISTICAL ANALYSIS?

A. Generally speaking, the goal of statistics is to analyze and interpret data and to objectively determine the reliability of the conclusions. These methods can be, and are, applied to almost every facet of everyday life. For example, one can look at the results of surveys to test and draw conclusions about public opinion, or at economic data to draw conclusions about the effects of an event or decision, or at the results of a scientific experiment to test whether a theory is correct. Or one can flip a coin to test whether it is fair. Statistical methods are often used to determine whether some factor, or factors, cause two populations to differ. For example a school system may use statistical methods to determine whether a given teacher produces a population of students that perform better on standardized tests than a population of students taught by other teachers. Each of these examples uses observed data to test a hypothesis and to draw conclusions about a population or populations.

To draw valid conclusions, one must address the reality that individual observations of data are often subject to random variation. On average, a fair coin will come up "heads" 50 percent of the time and "tails" the other 50 percent. If you could flip the coin an infinite number of times, you would tend to see results that are closer to an equal number of heads and tails. But in real world settings we do not have an infinite amount of data. If you flip a coin only twice, there is a 50 percent chance that it will come up one head and one tail, but there is also a 50 percent chance that it will produce either two heads or two tails. The coin may still be fair if it produces two heads or two tails; it's just that you only looked at two flips of the coin instead of a very large number. Thus, before jumping to a conclusion that the coin is unfair because it came up heads two

straight times, you need to consider the impact of random variation. That is also why survey results generally come with a margin of error. Statistical analysis provides a scientific method for factoring that random variation into the thought process.

Q. HOW DO STATISTICAL PRINCIPLES APPLY TO PERFORMANCE MEASUREMENT, STANDARDS, AND REMEDIES?

A. Performance measurement is just one of the many applications in which observed data is used to test a hypothesis about the population. In the case of performance measurement, Ameritech Illinois reports data about its performance of numerous functions. The data are reported separately for each CLEC, for a number of different products or services, for designated geographic areas. Ameritech Illinois compares each of these performance results to a standard. These standards are based on the principle of nondiscrimination or "parity" that is set forth in the 1996 Telecommunications Act, and in the criteria established by this Commission and the FCC. In most cases, the standard is "parity" between the wholesale function and a retail analog. In some cases, there is no retail analog, and the standard is set by a "benchmark": for example, 99 percent of mechanized completions are to be returned within 1 hour.

You can think of each of these performance measurements as a "test," the purpose of which is to draw a conclusion about whether Ameritech Illinois is satisfying its obligation to provide nondiscriminatory service, and whether Ameritech Illinois' personnel, electronic systems, and procedures are functioning in a nondiscriminatory manner. As Mr. Fioretti describes in his affidavit, there are over 160 performance measures, which are further divided into thousands of product, service, and geographic categories.

As with data used in other tests, performance data are subject to random variation. For example, in December 2000 (simulated data), the average time required for Ameritech Illinois to install retail residential POTS not requiring a field visit in the Chicago geographic region service was 0.65 days. But obviously each and every installation would not take exactly 0.65 days. Instead, some installations take less time, while others take more. Thus, if you randomly picked some installations out of the monthly total, the average time for those installations would likely be different from the overall average. Similarly, if you looked at the average installation time for CLEC customers, it too would likely be different from the overall retail average, even though CLEC customers are receiving the same level of service as Ameritech's own retail customers.

Q. WHAT CAUSES RANDOM VARIATION IN PERFORMANCE DATA?

A. There are an almost infinite number of causes. For example, one installation might take longer than another because of weather, or traffic conditions, or because the installation itself is more or less complicated or difficult to complete. These random events will cause the observed level of service provided to the CLEC's customers to appear better than that provided to Ameritech Illinois customers in some months and worse in other months, simply due to indiscriminant random variation.

Q. WHY IS IT IMPORTANT TO ADDRESS SUCH RANDOM VARIATION IN A PERFORMANCE REMEDY PLAN?

A. A remedy plan is designed to enforce performance standards and the underlying requirement of nondiscrimination. The basic idea is that, if Ameritech Illinois is discriminating against a CLEC or CLECs, it will pay a remedy to that CLEC (or to the

State). This provides compensation to the affected parties, and it provides an incentive to Ameritech Illinois to behave in a nondiscriminatory fashion.

If a remedy plan is to really serve its purpose, it should require remedies only where discrimination has really occurred. As I said earlier, individual performance observations (and thus average performance for different groups or samples of observations) are subject to random variation. Thus, even if there is no discrimination, you will see a difference in the average performance between two groups of performance metrics data. In fact, as I noted above, you will see a difference in performance between two randomly selected groups taken entirely from Ameritech Illinois retail data, even though Ameritech Illinois by definition does not discriminate against itself.

A remedy plan that forces Ameritech Illinois to pay remedies when it does *not* discriminate will not create the proper incentives. It would be like making parking meters with random timers on them and then giving motorists parking tickets when the meters randomly expired. It is simply an arbitrary and capricious transfer of funds from one company to another that provides no benefit to consumers or enhancement to competition. The payment of remedies when the underlying level of service is in parity reduces Ameritech Illinois' incentive to provide parity of service. Furthermore, as discussed in more detail below, remedy payments when parity exists will inhibit effective competition in the market and reduce the incentive for Ameritech Illinois to introduce new products and technology. The purpose of statistical analysis is to account for random variation and thus increase the chance that when remedies are paid they result from actual disparity in service.

Q. HOW DO THE COMPETING REMEDY PLANS ADDRESS THIS ISSUE?

A. The Ameritech Illinois and CLEC remedy plans both recognize the problem of random variation, but they take very different approaches to address it. In Section II, I describe the statistical methodology that Ameritech Illinois uses in its current remedy plan. In Section III, I contrast the partially developed methodology that the CLECs have developed to date, and show why Ameritech Illinois' plan is preferable. Section IV illustrates how the respective remedy plans work, by using simulated performance data for a three-month period.

II. DESCRIPTION OF STATISTICAL METHODOLOGY IN AMERITECH ILLINOIS' REMEDY PLAN.

Q. PLEASE OUTLINE THE APPROACH THAT AMERITECH ILLINOIS TAKES IN ITS REMEDY PLAN TO ADDRESS RANDOM VARIATION.

A. Ameritech Illinois' remedy plan follows the general approach the CLECs originally developed. The test that the CLECs proposed was a "Z-test."[1]

The basic idea of the Z-test is to consider the size of the difference between observed performance and the applicable standard to determine whether the difference is larger than what would often be found due to simple random variation in the data. The larger the difference, the more likely it is that there is some underlying disparity in performance as opposed to some observed difference that happened by random chance. For example, if retail repairs take 24 hours, one is more likely to find a real disparity in

---

1 The test for interval measures is modified from the standard Z-test found in statistical texts. Throughout my testimony, I will generically refer to both the modified Z-test and the standard Z-test as the "Z-test."

performance if wholesale repairs take 240 hours than if wholesale repairs take 24.001 hours. Or, one is more likely to decide that a coin is unfair if it turns up "heads" 50 out of 50 times, than if it comes up heads 26 out of 50 times. Statistical science provides a rigorous way of looking at the difference, and the extent of the random variation in the data to figure out how likely it is that the difference is due to some real disparity as opposed to random chance – in other words, that the difference is "statistically significant." Ameritech Illinois' plan uses tests designed to achieve a 95 percent confidence level; roughly speaking, this means that an apparent shortfall in performance is considered "statistically significant" if the odds are 95 percent that it is caused by behavior rather than random chance.

Of course, that means the odds are still 5 percent that the apparent shortfall is caused by random chance, and with a large number of performance tests, a 5 percent error rate virtually guarantees that some measures will appear disparate even when they are in parity. With the introduction of remedy payments, this means that Ameritech Illinois would pay remedies on 5 percent of all tests performed even when the level of service to CLEC customers and Ameritech Illinois retail customers are in perfect parity. Therefore, as an integral part of achieving the CLECs' 95 percent confidence level, the CLECs proposed (and Ameritech Illinois has adopted) a standard statistical technique that considers the results of all the individual tests in the aggregate. The principle here is that, at the aggregate level, one is more likely to find a real disparity in performance if 100 out of 100 individual tests suggest a "failure" or disparity, than if only 1 out of 100 tests suggests disparity. In fact, given a 5 percent error rate, you would expect a few individual tests to suggest disparity based solely on random variation rather than any real

disparity in behavior. Again, statistical analysis provides a scientific method for determining how many individual test "failures" are expected due to random chance, and how many failures would be necessary to suggest a real disparity. The number of such failures is called "k" and the method for calculating that number is the "K table."

This overall integrated method of testing for differences between groups across a large number of tests is based on statistical techniques that are commonly accepted in the field of statistical research. They were developed for use in performance testing by the CLECs, and they have been approved by state commissions and the FCC. It is these very tests that form the basis of the Ameritech Illinois remedy plan.

Q. PLEASE TELL US IN MORE DETAIL ABOUT THE INDIVIDUAL PERFORMANCE TESTS AMERITECH ILLINOIS USES.

A. As I said earlier, most of Ameritech Illinois' wholesale performance standards are based on a comparison to retail performance. For example, one measure compares the average time to repair the CLEC's resale residential POTS service to the average repair interval for Ameritech Illinois' retail residential POTS service. I call these tests "parity" tests. In performance testing, we are checking to see if there is disparity between retail and wholesale performance for the month. (I describe the other kinds of tests, "benchmark" tests, later on in my testimony.) More specifically, we are trying to determine with a reasonable degree of scientific certainty whether retail performance is *better* than wholesale performance, because we are trying to guard against the possibility that Ameritech Illinois would treat itself better than it treats CLECs. The remedy plan is not concerned with situations in which retail performance is worse than wholesale, and Ameritech Illinois does not receive a credit in those situations. Because we are

concerned with potential disparities only in one direction (worse than retail performance), we describe the test as a one-tailed test.

Q. HOW DOES AMERITECH ILLINOIS TEST PARITY BETWEEN RETAIL AND WHOLESALE PERFORMANCE?

A. As I described above, the objective is to look at the difference between retail and wholesale performance, and figure out whether the difference falls within the range we would expect due to mere random variation, or whether it is large enough to suggest some underlying disparity in performance. For example, in December 2000 (simulated data) it took Ameritech Illinois 0.45 days, on average, to install wholesale residential POTS without fieldwork in Chicago. There were 8,792 such installations; some took more time than average, some less. The average interval for one CLEC (code-named "146" to protect its identity) that month was 0.55 days; there were 496 installations, some longer, some shorter than average, the same as with retail. Because the individual observations are different, the difference between the retail and wholesale averages may simply be the product of random chance: which carrier happened to get more of the shorter installations that month.

Q. HOW DO YOU DECIDE WHETHER THE DIFFERENCE IS DUE TO RANDOM CHANCE?

A. Just as there are commonly accepted ways to measure distance (feet, miles, meters) or time (minutes, hours, etc.), statistical science provides a number of standard ways to measure the degree of variation in data and to estimate the range of random variation we would expect between two samples of such data. Here, we look at the degree of variation in the retail data to determine the amount of variation that would be expected due to

random chance. Differences between retail observations are, almost by definition, due to random variation rather than discrimination because Ameritech Illinois would not discriminate against itself. Where there are at least 30 CLEC observations for interval measures, we measure the degree of variation by using a statistical measure called the "modified Z statistic." (As I describe below, there is an alternative statistical measure that works better when there are less than 30 observations.)

1. Parity Tests: Sample Size Of 30 Or More

Q. WHAT IS THE SOURCE OF THE MODIFIED Z STATISTIC?

A. It was developed and proposed by a group of carriers – AT&T, MCI (now WorldCom), Sprint, and LCI – that called themselves the Local Competition User Group or "LCUG." They first advanced the modified Z-statistic in 1998, during the FCC's rulemaking on performance measurement. Over time, a consensus developed as other carriers agreed to LCUG's approach. Ameritech Illinois accepted the modified Z test in the spirit of compromise, in order to allow statistical testing to commence.

Q. WHY IS IT CALLED A "MODIFIED" Z STATISTIC?

A. The modified Z-statistic is based on a commonly used statistical measure known as the "Z-statistic," which is designed to assess whether an observed difference between two averages is statistically significant. The LCUG worried that the standard Z-test, while testing for differences in the average service provided to CLEC and Ameritech Illinois retail customers, did not test for differences in the consistency or "variance" of service provided to CLEC customers. The LCUG suggested that incumbent carriers would provide their own retail customers and CLEC customers with the same average service, but could achieve this same average level of service to the CLEC customers by providing

some CLEC customers with very high quality service and other CLEC customers with very low quality or slow service. The LCUG suggested that this type of increased variation in level of service to CLEC customers would not only constitute a lower quality of service in itself, but that it would also reduce the effectiveness of a Z-test to detect any potential differences in the average level of service provided to CLEC and Ameritech Illinois customers. The LCUG, therefore, proposed the "modified" Z-test, which substitutes the incumbent LEC ("ILEC") standard deviation for the CLEC standard deviation in the standard Z formula. The benefit of this test is that differences in the variation in service provided to CLEC customers would not reduce the effectiveness of the test in detecting differences in the average level of service provided to CLEC and Ameritech Illinois customers.

For metrics that are measured as rates and proportions, the average performance determines the variance. It is not possible to maintain the same average performance while increasing the variance in performance of a rate or a proportion. Therefore for rates and proportions there is no need to modify the standard Z-test.

Q. WHY IS THE MODIFIED Z TEST NOT USED FOR PROPORTIONS AND RATES?

A. It is not necessary. As I stated above, the reason for using the modified Z-test for intervals is because there is concern that the ILEC would have an incentive to inflate the variability of the CLEC performance in order to pass the parity test while still providing better service to itself. For rates and proportions, the ILEC is unable to pursue a strategy of increasing the variance of the data in order to achieve a lower Z-score, because for rates and proportions the variability (variance) cannot be increased without also altering

the proportion or rate itself. Therefore, the more conventional pooled Z-test is used for these measures.

Q. HOW ARE THE Z-STATISTICS CALCULATED?

A. The Z-statistic is a commonly accepted statistical tool that uses the mean of the data and a commonly accepted measure of variation called the "standard deviation," which measures the normal or "standard" amount by which the individual data observations differ or "deviate" from the overall average. The Greek letter sigma (σ) is a shorthand symbol for the standard deviation. There are slightly different formulas for computing the Z-statistic, depending on whether the performance measurement is an average (the average time to repair service), a percentage (the percentage of due dates missed), or a rate (the rate of trouble reports). These formulas are well known in the field, and they are illustrated in section 3.0 of the Ameritech Illinois remedy plan. Attachment B to my testimony illustrates the calculation of a modified Z-statistic with a numerical example.

Q. AFTER CALCULATING "Z," WHAT IS THE NEXT STEP IN TESTING PARITY?

A. The next step is to see if Z, the measure of difference between average wholesale and retail performance, falls within the range of differences we would expect due to random chance. We do that by comparing the modified Z-statistic to the amount of difference one would expect from random variation, which is called the "critical" Z value.

Q. HOW DOES AMERITECH ILLINOIS CALCULATE THE CRITICAL Z VALUE?

A. Using standard statistical methods, AT&T developed a table that lists the combination of critical *Z*-values, and the number of apparently disparate test results, that would indicate true underlying disparity at a 95 percent confidence level. In other words, if the test indicates a failure or disparity, there is 95 percent confidence that there was a real disparity. Conversely, you can say that there is a 5 percent risk that the test will indicate disparity (in error) where there is none. These false alarms are referred to as "Type I" errors.

To determine the relevant critical Z-value under the Ameritech Illinois plan, all you need to know is the number of parity tests that are performed for a given CLEC. Once this is known you simply look up the combination of the critical Z-value and the number of missed parity tests needed to demonstrate disparity at a 95 percent confidence level.

Q. WHY DOES AMERITECH ILLINOIS USE A 95 PERCENT CONFIDENCE LEVEL?

A. The 95 percent confidence level is commonly used often in the field of statistical science. A 99 percent confidence level would have been a valid choice as well: It is also used frequently in scientific and statistical research. AT&T first suggested using the 95 percent confidence level for performance testing in 1998, during the FCC rulemaking on performance measurement.

As with the modified Z-test, other carriers quickly formed a consensus that 95 percent confidence would be appropriate. Although Ameritech Illinois believes that the 99 percent confidence interval would be appropriate, Ameritech Illinois accepted the 95

percent confidence interval in the spirit of compromise even though it would lead to an increased number of false findings of disparity. The FCC then approved the 95 percent confidence level when it approved the application of Bell Atlantic (now Verizon) to provide long-distance service in New York.

The 95 percent confidence interval has also been adopted by KPMG Consulting, which is conducting statistical tests of performance as part of its independent audit of OSS in Illinois and throughout the region.

Q. HOW DOES THE TABLE OF CRITICAL Z-VALUES IN THE AMERITECH ILLINOIS REMEDY PLAN WORK?

A. The table that appears in Section 9.3 of the Ameritech Plan is reproduced below. Table 1 lists the Z-values and number of apparently disparate tests results ("k") that would be needed to demonstrate disparity at the 95 percent confidence interval. It is based on a table developed by AT&T.[2]

[2] Affidavit of Colin Mallows, CC Docket No. 98-56 (Attachment C), pp. 18-19. The k-table within the Ameritech Illinois plan differs slightly from the one developed by AT&T. It is my understanding that Ameritech Illinois would be willing to alter the k-table in the Remedy Plan to make it consistent with the one developed by AT&T.

Table 1. Critical Z - Statistic Table

Number of Performance Tests	K Values	Critical Z-value
1	0	1.65
2	0	1.96
3	0	2.12
4	0	2.23
5	0	2.32
6	0	2.39
7	0	2.44
8	1	1.69
9	1	1.74
10-19	1	1.79
20-29	2	1.73
30-39	3	1.68
40-49	3	1.81
50-59	4	1.75
60-69	5	1.7
70 –79	6	1.68
80 – 89	6	1.74
90 – 99	7	1.71
100 – 109	8	1.68
110 –119	9	1.7
120 – 139	10	1.72
140 – 159	12	1.68
160 – 179	13	1.69
180 – 199	14	1.7
200 – 249	17	1.7
250 – 299	20	1.7
300 – 399	26	1.7
400 – 499	32	1.7
500 – 599	38	1.72
600 – 699	44	1.72
700 – 799	49	1.73
800 – 899	55	1.75
900 – 999	60	1.77
1000 and above	Calculated for Type-1 Error Probability of 5 percent	Calculated for Type-1 Error Probability of 5 percent

The first column of Table 1, lists the possible numbers of performance tests, while the third column lists the critical Z-value that applies to that number of tests. All you need to do is take the number of performance tests for the CLEC in question with at least 10 observations, find the applicable row in the table, then go over to the "Critical Z" column and find the applicable value. For example, if 19 performance tests apply to CLEC "1", the critical Z-value would be 1.79. If the number of performance tests for CLEC "2" is between 600 and 699, the critical Z-value would be 1.72. The math is based on standard statistical formulas (used by AT&T), and has already been done and recorded on the table.

After you take the critical Z- value from the table, you compare the actual Z-statistic for each performance test in question. If the Z-statistic is lower than the critical Z- value, the difference between wholesale and retail performance is not large enough to suggest disparity with 95 percent confidence, and we move on to the next test. The difference is not "statistically significant." If the number of recorded tests with Z- values larger than the critical Z exceeds the value of k in the third column of the same row, we can conclude with 95 percent confidence that there was disparity of service.

Q. CAN YOU PROVIDE SOME EXAMPLES OF HOW THE Z-TEST WORKS?

A. Yes. In December 2000 (from simulated data), the performance results for CLEC "174" showed activity in 76 performance measurement categories. Accordingly, based on the table of critical values at pages 11-12 of the remedy plan, the critical Z value for that CLEC was 1.68. The data for CLEC "174" included the following results for performance measure 27:

1) Mean Installation Interval - POTS - Residential - Fieldwork (Days), Chicago region

CLEC mean: 3.00 days ILEC mean: 4.87 days Z value: -6.09 Critical Z: 1.68

2) Mean Installation Interval - POTS - Business - No Fieldwork (Days), Chicago region

CLEC mean: 0.88 days ILEC mean: 0.30 days Z value: 1.56 Critical Z: 1.68

3) Mean Installation Interval - POTS – Residential - Feildwork (Days), IL North Central

CLEC mean: 4.00 days ILEC mean: 3.55 days Z value: 3.37 Critical Z: 1.68

For the first measure, the CLEC's results were better than retail. Because the remedy plan is only concerned with a disparity that goes *against* the CLEC, no further analysis is performed.

For the second measure, CLEC installations took more time than retail. The Z statistic, however, was only 1.56, less than the critical z value of 1.68. As a result, we conclude that the difference is due to random variation, and no further analysis is performed.

For the third measure, the Z-statistic of 3.37 exceeds the critical Z value of 1.68. This result suggests disparity, but with 95 percent confidence (or a 5 percent risk of error). Once all of the other parity tests for that CLEC have been performed in the month we can determine if the overall performance for that CLEC suggests that there is parity of service. Given that the CLEC has 76 performance metrics required for testing the k value drawn from Table 1 for this CLEC will be six. If more than six of these tests have a Z-

value exceeding 1.68 then we would be able to state that there is evidence of disparity based on a 95 percent confidence interval.

Q. WHAT IS THE PURPOSE OF THE THIRD COLUMN OF THE TABLE 1?

A. When thousands of statistical tests are performed, and each has a 5 percent Type I error rate, as is the case with performance testing, it is virtually guaranteed that large numbers of tests will appear to show disparity even when service is in perfect parity. The k-value in Table 1, in combination with the critical Z- value in that same table, establishes the number of apparent failures that would be needed to show a real disparity with the 95 percent confidence level the CLECs demanded.

Q. IS THE K VALUE EXACTLY 5 PERCENT OF THE NUMBER OF TESTS?

A. No. To say the individual statistical tests yield Type I errors 5 percent of the time *on average*, is the same thing as saying that the rate of flipping a fair coin and getting "heads" is 50 percent on average. For smaller sample sizes (say 30 flips), the actual number of heads will frequently vary quite a bit from that 50 percent average. We would expect that about half the time it will be higher and half the time lower. Similarly, given a Type I error rate of 5 percent, the number of false alarms for a given CLEC will likely exceed 5 percent half the time. Thus, setting "k" at exactly five percent is not statistically valid: It would give you only 50 percent "confidence" in the result.

To achieve the standard level of confidence, 95 percent, k is set slightly higher than 5 percent. As one would expect, where the number of measurement categories is small, the number of measures excluded is slightly higher than 5 percent: hence, eight categories would be excluded if 100 categories had data. But as the number of categories

increases – and there are now several thousand measurement categories – the value of k does approach 5 percent.

Q. IF THE K VALUE SAYS THAT SOME, BUT NOT ALL, OF THE APPARENT DISPARITIES ARE DUE TO RANDOM CHANCE, WHICH OF THOSE DISPARITIES ARE USED FOR ASSESSING REMEDIES?

A. As Mr. Fioretti shows in his affidavit, the remedy amounts for each performance tests reflect the importance of the related performance measure tested: either "high," "medium" or "low" priority. The K table is applied to the low priority measures first. Thus, Ameritech Illinois will pay on the highest-priority measures, which generally have the potential for highest remedies.

2. Parity Tests: Sample Sizes Less Than 30

Q. DOES AMERITECH ILLINOIS USE THE Z-TEST FOR ALL "PARITY" TESTS?

A. No. As I mentioned earlier, the Z-test is unlikely to function well on small sample sizes, such as less than 30 observations. Therefore, Ameritech Illinois uses alternative tests for sample sizes less than 30.

Q. WHY DOESN'T THE Z-TEST WORK ON SAMPLE SIZES UNDER 30?

A. Typically, as you look at more and more individual pieces of data or observations, they tend to fall into a pattern or distribution. Experience has shown that frequently once you look at 30 observations, the distribution of averages drawn from the population should be "normal," a statistical term that means the distribution is bell-shaped with about two-thirds of all observations falling within one standard deviation of the average. (This

tendency of the sample averages toward normality results from what statisticians call the "Central Limit Theorem.") The Z-test is designed to work well on these normal distributions. But if there are fewer than 30 observations, the sample averages are often fairly different from normal. And in this situation other statistical tests tend to be more effective and appropriate to use. For example, if you randomly pick only five repairs out of the universe of all repairs, you might end up picking the rare cases that are far away from the average (say, the repair that took twice as long as normal due to unusual weather or traffic), and thus do not fairly reflect the population of repairs as a whole. In those cases, the critical Z-value may not reflect the true cut-off for the 95 percent confidence interval.

Q. HOW DOES AMERITECH ILLINOIS ADDRESS THESE SITUATIONS?

A. The remedy plan uses alternative statistical tests known as "permutation tests," which are not sensitive to the non-normal distributions of the sample means that are often seen in sample sizes of less than 30 observations. In fact, permutation tests are always at least as good as Z-tests no matter how large the sample size; however, in large sample sizes they require a large number of repetitive calculations. Because the greatest advantage of the permutation test compared to the Z-test is observed when distributions are non-normal, Ameritech Illinois agreed with the consensus that formed around the practice of using permutation tests only when sample sizes were less than 30.

3. "Benchmark" Tests

Q. THE Z-TEST AND PERMUTATION TESTS ARE APPLIED TO "PARITY" TESTS. DOES AMERITECH ILLINOIS ADDRESS RANDOM VARIATION IN "BENCHMARK" TESTS?

A. Yes. In a benchmark test, wholesale performance is compared to a specified target. While the target itself doesn't move, wholesale performance is still subject to random variation, whatever you choose to compare against it.

Q. HOW DOES RANDOM VARIATION AFFECT BENCHMARK TESTS?

A. Let's use the return of mechanized completions as an example. The benchmark for this measure is 99 percent returned within 1 hour. Not every completion takes exactly 1 hour. And even if Ameritech Illinois returns 99 percent within 1 hour for all CLECs on average, you would expect the results for individual CLECs to vary: for half the CLECs, performance would be higher than 99 percent, but for the other half, performance would be less than 99 percent. Ameritech Illinois does not get credit for the half that are a little higher than the benchmark, and it should not be penalized for the half that are a little lower.

Q. HOW DOES AMERITECH ILLINOIS ADDRESS RANDOM VARIATION IN BENCHMARK TESTS?

A. The method is the same as that used by Southwestern Bell, and approved by the Texas commission and the FCC, in Texas. It does not use statistical analysis. It is my understanding that there is a buffer calculated for the benchmark as a compromise in Texas to reflect the random variations that occurs with all performance metrics.

B. <u>Analysis Of Disparity</u>

Q. IF THE STATISTICAL TESTS YOU DESCRIBE SHOW A DISPARITY, DOES THAT PROVE DISCRIMINATION BY AMERITECH ILLINOIS?

A. Not at all. First, as I described above, the statistical tests are designed to achieve only 95 percent confidence that disparity exists even when the statistical tests indicate there is disparity.

More importantly, these statistical analyses only indicate a *numerical* discrepancy between wholesale performance and the applicable standard (either retail performance or the benchmark). They do not identify the cause of that discrepancy. The discrepancy might have been the fault of Ameritech Illinois, but it might also be the result of something outside of Ameritech Illinois' control, such as an Act of God, or some mistake or misconduct by the CLEC or a third party (for example, a technician hired by the CLEC to coordinate a loop cut-over with Ameritech Illinois).

Q. CAN YOU GIVE US SOME EXAMPLES OF DISCREPANCIES CAUSED BY FACTORS OUTSIDE OF AMERITECH ILLINOIS' CONTROL?

A. Certainly. Let's assume that Ameritech Illinois and a CLEC each experience 100 "trouble reports" in January, and that the mean time to repair for the Ameritech Illinois customers was 3.4 hours, while the CLEC's customers experienced an average time of 5.0 hours. Statistical analysis might conclude there was a disparity, with 95 percent confidence. But further analysis might show that the time to restore service was exactly 7 hours during the first week (due to a blizzard) for all carriers, and 3 hours the rest of the month (again, for all carriers). It might also show that 50 percent of the CLEC customers reported trouble during the blizzard week, while only 10 percent of the Ameritech Illinois customers reported trouble that week. Thus, the CLEC average would be 5 hours (50 percent at 3 hours, and 50 percent at 7 hours) while the Ameritech Illinois average would be 3.4 hours (90 percent at 3 hours, and 10 percent at 7 hours). The discrepancy would

not be due to any wrongdoing by Ameritech Illinois, but would result from the fact that a higher percentage of CLEC repairs happened to coincide with the blizzard.

Q. HOW DOES THE AMERITECH ILLINOIS REMEDY PLAN ADDRESS THESE POSSIBILITIES?

A. The remedy plan takes a pro-CLEC approach. It presumes that the discrepancy was the fault of Ameritech Illinois and requires Ameritech Illinois to pay a remedy unless the Commission finds that a remedy is not warranted. The plan then establishes an expedited procedure for Ameritech Illinois to seek a waiver from the Commission. If Ameritech Illinois does not initiate that procedure before the date remedy payments are due, the pro-CLEC presumption stands and Ameritech Illinois must pay the remedy. The waiver procedure is described in more detail in the affidavit of Mr. Fioretti.

Q. IF THE STATISTICAL TESTS YOU DESCRIBE SHOW DISPARITY DOES THAT MEAN THAT DIFFERENCES IN SERVICE ARE LARGE OR NOTICABLE TO CUSTOMERS?

A. No. These statistical tests are designed to determine if there is enough empirical evidence to establish some disparity no matter how large or small the disparity. A one-minute disparity in installing POTS between wholesale and retail, for example, would probably not affect consumer decisions or even noticed by consumers. The tests used by Ameritech Illinois do test for these small differences and assess remedies based on them, but they by no means imply that these differences are large enough to be relevant to consumers or competition. In this way, Ameritech Illinois' statistical tests and remedy plan are pro-CLEC, paying remedies in cases where there is a statistical difference even if the difference is too small to be relevant to consumers.